THEORY OF PROBABILITY

BY

HAROLD JEFFREYS
FORMERLY PLUMIAN PROFESSOR OF ASTRONOMY
UNIVERSITY OF CAMBRIDGE

THIRD EDITION

CLARENDON PRESS · OXFORD

Oxford University Press, Walton Street, Oxford OX2 6DP
Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland
and associated companies in
Beirut Berlin Ibadan Nicosia

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

First published in the International Series
of Monographs on Physics 1939
Second edition 1948
Third edition 1961
First issued in paperback (with corrections) 1983
Reprinted 1985

British Library Cataloguing in Publication Data
Jeffreys, Harold
Theory of probability.—3rd ed.— (the International series of monographs on physics)
1. Probabilities
I. Title II. Series
519.2 QA273
ISBN 0–19–853193–1

Printed and bound in Great Britain by
Biddles Ltd, Guildford and King's Lynn

PREFACE TO THE CORRECTED IMPRESSION

SOME corrections and amplifications have been made in the present version and an appendix on harmonic analysis and autocorrelation has been added.

I am indebted for helpful discussions to Professor D. V. Lindley, Professor H. E. Daniels, and Dr. A. M. Walker.

H. J.

Cambridge, 1966

hypotheses t … servational
data are pre … bability is
not to excee … t if a set of
positive qua … less than a
fixed quantit … words they
are the term … not directly
say very mu … parameter
capable of a … numerable.
We could, h … f intervals,
and the con … ler limiting
behaviour w … . Similarly,
if the law co … tinuous set
we could rec … th repeated
verification … rrect value
of the param … arbitrarily
short.

This argument covers most problems of estimation, but does not do all that is needed. Return to the problem of the falling body (1.0). The law is in the form

$$s = a+ut+\tfrac{1}{2}gt^2. \tag{1}$$

Here t and s are measured, a, u, g are parameters (that is, quantities common to every observation). a, u, g are *adjustable*; that is, their values are initially unknown, and they are to be determined as well as possible from the observations. If this was all, the above argument would qualitatively cover the ground, though we shall make it more precise later. This would be an *estimation problem.*

But we might consider the hypothesis

$$s = a+ut+\tfrac{1}{2}gt^2+a_3t^3+...+a_nt^n, \tag{2}$$

where n is greater than the number of observations and all coefficients are adjustable. For any set of observations the solution is wholly indeterminate, and gives no information at all about values of s at times other than those observed. Even for values of n equal to or smaller than the number of observations the uncertainty of each term will exceed the whole variation of s. But it would be preposterous to say on this ground that the observations give no information at intermediate times, when the first three terms, with suitable values of a, u, and g, in fact account for nearly all the variation of s at the observed times. The conclusion is that including too many terms will *lose* accuracy in prediction instead of gaining it. Thus we have the problem, given a set of measures: what set of coefficients in (2) should be taken as

adjustable (here not zero) in order to achieve the most accurate predictions? We certainly must take *some* as not adjustable; (1) corresponds to taking all of a_3, a_4,..., a_n as zero, and if any of them is taken as adjustable the result can be regarded as a different law. Then our problem is to assess probabilities of the different laws. These (if n is allowed to be arbitrarily large) constitute an enumerable set, and the prior probability that any one is right can be taken positive, subject to the condition of convergence. For any particular one the adjustable parameters can have a continuous probability distribution. Then the theory will lead to posterior probabilities for the various laws. This procedure constitutes a *significance test.*

Precise statement of the prior probabilities of the laws in accordance with the condition of convergence requires that they should actually be put in an order of decreasing prior probability. But this corresponds to actual scientific procedure. A physicist would test first whether the whole variation is random as against the existence of a linear trend; then a linear law against a quadratic one, then proceeding in order of increasing complexity. All we have to say is that the simpler laws have the greater prior probabilities. This is what Wrinch and I called the *simplicity postulate.* To make the order definite, however, requires a numerical rule for assessing the complexity of a law. In the case of laws expressible by differential equations this is easy. We could define the complexity of a differential equation, cleared of roots and fractions, by the sum of the order, the degree, and the absolute values of the coefficients. Thus

$$s = a$$

would be written as $$ds/dt = 0$$

with complexity $$1+1+1 = 3.$$

$$s = a+ut+\tfrac{1}{2}gt^2$$

would become $$d^2s/dt^2 = 0$$

with complexity $$2+1+1 = 4;$$

and so on. Prior probability 2^{-m} or $6/\pi^2m^2$ could be attached to the disjunction of all laws of complexity m and distributed uniformly among them. This does not cover all cases, but there is no reason to suppose the general problem insoluble. Detailed solutions on these lines of some of the more important problems of statistics are given in Chapters V and VI.

All the laws of classical physics are in fact expressible by differential equations, and those of quantum physics are derived from them by various systematic modifications. So this choice does take account of